

AUDIO FINGERPRINTING SYSTEM AND METHOD

5 FIELD OF THE INVENTION

The present invention is generally related to automatically identifying unknown audio pieces, and more specifically, to a system and method for efficiently identifying unknown audio pieces via their audio fingerprints.

BACKGROUND OF THE INVENTION

It is often desirable to automatically identify an audio piece by analyzing the content of its audio signal, especially when no descriptive data is associated with the audio piece. Prior art fingerprinting systems generally allow recognition of audio pieces based on arbitrary portions of the piece. The fingerprints in the fingerprint database are often time-indexed to allow appropriate alignment of a fingerprint generated based on the arbitrary portion with a stored fingerprint. Time-based fingerprinting systems therefore add an additional complicating step of locating a correct segment in the fingerprint database before any comparison may be performed.

The generating and storing of time-indexed audio fingerprints are redundant if an assumption may be made as to the portion of the audio piece that will be available for fingerprinting. For example, if it is known that the audio piece to be identified will always be available from the beginning of the piece, it is not necessary to maintain time-indexed fingerprints of the audio piece for the various time slices, nor is it necessary to time-align a query fingerprint with a stored fingerprint.

Another problem encountered in prior art fingerprinting systems is that as the number of registered fingerprints in the fingerprint database increases, the time expended to obtain a match also increases.

Thus, what is needed is a fingerprinting system that provides a reliable, fast, and robust identification of audio pieces. Such a system should be configured to reduce the search space in performing the identification for a better matching accuracy and speed.

SUMMARY OF THE INVENTION

According to one embodiment, the invention is directed to a method for making choices from a plurality of audio pieces where the method includes: receiving an audio fingerprint of a first audio piece; searching a database for the audio fingerprint; retrieving an audio profile vector associated with the audio fingerprint, the audio profile vector quantifying a plurality of attributes associated with the audio piece; updating user preference information based on the audio profile vector; and selecting a second audio piece based on the user preference information.

According to another embodiment, the invention is directed to an audio fingerprinting method that includes: receiving an audio signal associated with an audio piece; obtaining a plurality of frequency measurements of the audio signal; building a matrix A based on the frequency measurements; performing a singular value decomposition on the matrix A , wherein $A = USV^T$; retrieving one or more rows of matrix V^T ; associating the retrieved rows of matrix V^T with the

audio piece; and storing the retrieved rows of matrix V^T in a data store.

5 According to another embodiment, the invention is directed to an audio indexing method that includes: receiving an audio signal of an audio piece; automatically obtaining from the audio signal a list of musical notes included in the audio piece; determining from the audio signal a prominence of
10 the musical notes in the audio piece; selecting a pre-determined number of most prominent musical notes in the audio piece; generating an index based on the selected musical notes; and searching a database based on the generated index.

15 According to another embodiment, the invention is directed to a method for generating an identifier for an audio class where the method includes: selecting a plurality of audio pieces associated with the audio class; computing an audio fingerprint for each selected audio piece; calculating
20 an average of the computed audio fingerprints; generating an average fingerprint based on the calculation; associating the average fingerprint to the audio class; and storing the average fingerprint in a data store.

25 According to another embodiment, the invention is directed to an audio selection system that includes: a first data store storing a plurality of audio fingerprints for a plurality of audio pieces; a second data store storing a plurality of audio profile vectors for the plurality of audio
30 fingerprints, each audio profile vector quantifying a plurality of attributes associated with the audio piece corresponding to the audio fingerprint; means for searching the first data store for an audio fingerprint of a first audio
35 piece; means for retrieving from the second data store an

audio profile vector associated with the audio fingerprint;
 means for updating user preference information based on the
 5 retrieved audio profile vector; and means for selecting a
 second audio piece based on the user preference information.

According to another embodiment, the invention is
 directed to an audio fingerprinting system that includes a
 10 processor configured to: receive an audio signal associated
 with an audio piece; obtain a plurality of frequency
 measurements of the audio signal; build a matrix A based on
 the frequency measurements; perform a singular value
 decomposition on the matrix A , wherein $A = USV^T$; retrieve one
 15 or more rows of matrix V^T ; and associate the retrieved rows of
 matrix V^T with the audio piece. The audio fingerprint system
 also includes a data store coupled to the processor for
 storing the retrieved rows of matrix V^T .

20 According to another embodiment, the invention is
 directed to an audio indexing system that includes a means for
 receiving an audio signal of an audio piece; means for
 automatically obtaining from the audio signal a list of
 musical notes included in the audio piece; means for
 25 determining from the audio signal a prominence of the musical
 notes in the audio piece; means for selecting a pre-determined
 number of most prominent musical notes in the audio piece;
 means for generating an index based on the selected musical
 30 notes; and means for searching a database based on the
 generated index.

According to another embodiment, the invention is
 directed to a system for generating an identifier for an audio
 class where the system includes: means for computing an audio
 35 fingerprint for each of a plurality of selected audio pieces;

1 **50659/JEC/P396**

means for calculating an average of the computed audio fingerprints; means for associating the calculated average to
5 the audio class; and means for storing the calculated average in a data store.

According to another embodiment, the invention is directed to an article of manufacture comprising a computer
10 readable medium having computer usable program code containing executable instructions that, when executed, cause a computer to perform the steps of: obtaining a plurality of frequency measurements of an audio signal associated with an audio
15 piece; building a matrix A based on the frequency measurements; performing a singular value decomposition on the matrix A , wherein $A = USV^T$; retrieving one or more rows of matrix V^T ; associating the retrieved rows of matrix V^T with the
20 audio piece; and storing the retrieved rows of matrix V^T in a data store.

According to another embodiment, the invention is directed to an article of manufacture comprising a computer
readable medium having computer usable program code containing
25 executable instructions that, when executed, cause a computer to perform the steps of: automatically obtaining from an audio signal of an audio piece, a list of musical notes included in the audio piece; determining from the audio signal a
prominence of the musical notes in the audio piece; selecting
30 a pre-determined number of most prominent musical notes in the audio piece; generating an index based on the selected musical notes; and searching a database based on the generated index.

These and other features, aspects and advantages of the
35 present invention will be more fully understood when considered with respect to the following detailed description,

appended claims, and accompanying drawings. Of course, the actual scope of the invention is defined by the appended
5 claims.

BRIEF DESCRIPTION OF THE DRAWINGS

FIG. 1 is a schematic block diagram of an audio fingerprinting system according to one embodiment of the
10 invention;

FIG. 2 is a flow diagram of a process for generating an audio fingerprint according to one embodiment of the invention;

15 FIG. 3 is a flow diagram of a process for analyzing an extracted audio fingerprint for a match against registered fingerprints according to one embodiment of the invention;

FIG. 4 is a flow diagram of a process for analyzing an extracted audio fingerprint for a match against registered fingerprints according to an alternative embodiment of the invention;

20 FIG. 5 is a flow diagram of a process for assigning a database index to an audio piece according to one embodiment of the invention;

FIG. 6 is a flow diagram of a process for generating an identifier for a particular musical class according to one embodiment of the invention; and

30 FIG. 7 is a schematic block diagram of a computer network with one or more devices utilizing the audio fingerprinting system of FIG. 1 according to one embodiment of the invention.

DETAILED DESCRIPTION OF THE INVENTION

FIG. 1 is a schematic block diagram of an audio fingerprinting system 10 according to one embodiment of the invention. The system includes an audio file reader 12 for reading different types of audio files 11 or an audio input, and for outputting wave (.wav), MP3 file, or the like. The audio file reader 12 may be, for example, a CD player, DVD player, hard drive, or the like. The file reader 12 may be coupled to an MP3 decoder 14 for decoding MP3 files output by the audio file reader 12. Other types of decoders may also be used for decoding other types of encoded audio files.

The audio file 11 provided to the audio file reader 12 may be an entire audio piece or a portion of the audio piece to be recognized or registered. According to one embodiment of the invention, the audio file contains at least the first thirty seconds of the audio piece. A person of skill in the art should recognize, however, that shorter or longer segments may also be used in alternative embodiments.

The received audio file 11 is transmitted to a music preprocessor 16 which, according to one embodiment of the invention, is configured to take certain pre-processing steps prior to analysis of the audio file. Exemplary pre-processing steps may include normalizing the audio signal to ensure that the maximum level in the signal is the same for all audio samples, transforming the audio data from stereo to mono, eliminating silent portions of the audio file, and the like. A person skilled in the art should recognize, however, that the pre-processing step may be eliminated or may include other types of audio pre-processing steps that are conventional in the art.

The preprocessor 16 is coupled to a fingerprint
 extraction engine 18, fingerprint analysis engine 20, indexing
 5 engine 22, and class identification engine 24. According to
 one embodiment of the invention, the engines are processors
 that implement instructions stored in memory. A person of
 skill in the art should recognize, however, that the engines
 10 may be implemented in hardware, firmware (e.g. ASIC), or a
 combination of hardware, firmware, and software.

 According to one embodiment of the invention, the
 fingerprint extraction engine 18 automatically generates a
 compact representation, hereinafter referred to as a
 15 fingerprint of signature, of the audio file 11, for use as a
 unique identifier of the audio piece. According to one
 embodiment of the invention, the audio fingerprint is
 represented as a matrix.

 The fingerprint analysis engine 20 analyzes an audio
 fingerprint generated by the fingerprint extraction engine 18
 for a match against registered fingerprints in a fingerprint
 database 26. Based on the match, either the fingerprint
 analysis engine or a separate engine coupled to the
 25 fingerprint analysis engine (not shown) retrieves additional
 data associated with the audio piece. The additional data may
 be, for example, an audio profile vector that describes the
 various attributes of the audio piece as is described in
 30 further detail in U.S. Patent Application Ser. No. 10/278,636,
 filed on October 23, 2002, the content of which is
 incorporated herein by reference. Of course, a person of
 skill in the art should recognize that other types of data may
 35 also be associated with the audio piece, such as, for example,
 title information, artist or group information, concert

information, new release information, and/or links, such as URL links, to further data.

5 The indexing engine 22 associates the extracted audio fingerprint with an index that may be used by the fingerprint analysis engine 20 to identify a subset of candidates in the fingerprint database 26. According to one embodiment of the invention, the index is generated based on the prominent
10 musical notes contained in the audio piece. Once the index is generated, a subset of audio fingerprints in the fingerprint database 26 to which the audio piece belongs may be identified.

15 The class identification engine 24 generates identifiers for different sets of audio pieces that belong to particular musical classes. According to one embodiment of the invention, the audio pieces in a particular musical class are
20 similar in terms of overall instrumentation/orchestration. For example, an exemplary musical class may be identified as including a jazz piano trio, acappella singing, acoustic guitar, acoustic piano, solo acoustic guitar with vocal, or the like. The various musical classes may then be included as
25 attributes of an audio profile vector where a values set for a particular musical class attribute indicates how close or far the audio piece is to the musical class. The identifiers and information about the various musical classes may then be
30 stored in a musical class database 28.

 The fingerprint database 26 stores a plurality of fingerprints of known audio pieces. The fingerprints may be grouped into discrete subsets based on the musical notes
35 contained in the audio pieces. Each audio fingerprint may be associated with the actual audio file, an audio profile

vector, a description of the audio piece (e.g. title, artist and/or group), concert information, new release information, URL links to additional data, and/or the like.

FIG. 2 is a flow diagram of a process for generating an audio fingerprint according to one embodiment of the invention. The process starts, and in step 100, the fingerprint extraction engine 18 or a separate fourier transform engine (not shown) calculates a Fast Fourier Transform (FFT) or the like, of the audio signal of the preprocessed audio piece for transforming the signal waveform in the time domain into a signal in the frequency domain. According to one embodiment of the invention, the FFT analysis is resampled to reduce the size of the data for subsequent processing.

Based on the FFT calculation, the fingerprint extraction engine 18 generates, in step 102, a TxF matrix A, where $T \geq F$. According to one embodiment of the invention, the rows of the matrix represent time, and the columns of the matrix represent frequency measurements, also referred to as bins, of the FFT.

In step 104, the fingerprint extraction engine 18 performs the well known matrix operation known as a Singular Value Decomposition (SVD) operation on matrix A. In general terms, SVD is a technique that reduces an original matrix into a product of three matrices as follows:

$$\text{SVD}(A) = USV^T$$

where U is a TxF orthogonal matrix, S is an FxF diagonal matrix with positive or zero valued elements, and V^T is the transpose of an FxF orthogonal matrix. According to one

embodiment of the invention, the rows of V transposed are the coordinates that capture the most variance, that is, retain the most information about the audio piece in decreasing order of significance as measured by the diagonal entries of the S matrix.

In step 106, the fingerprint extraction engine 18 extracts a predetermined number of rows from the matrix V^T and in step 108, builds a fingerprint matrix from the extracted rows. In step 110, the fingerprint matrix is set as the audio piece's fingerprint by associating the fingerprint matrix to the audio piece in any manner that may be conventional in the art.

In step 112, the fingerprint matrix is stored in a data store. The data store is the fingerprint database 26 if the fingerprint extraction is done for registration purposes. Otherwise, the data store is a temporary storage location for storing the fingerprint matrix for later retrieval by the fingerprint analysis engine 20 for comparing against registered fingerprints.

Unlike many audio fingerprints generated by prior art systems, the audio fingerprint generated via the SVD operation has no notion of time associated with it. A person of skill in the art should recognize, however, that time may be associated with the audio fingerprint generated via the SVD operation. In other words, the process of generating audio fingerprints described with relation to FIG. 2 may be extended to a time-based audio fingerprint system by assigning a time-index to the audio fingerprint generated via the SVD operation, and repeating the process for a moving window across the entire song.

According to one embodiment of the invention, the fingerprint extraction engine 18 may also incorporate prior art fingerprinting techniques such as, for example, spectral centroid and/or spectral flatness measures which result in time-indexed fingerprint measurements. If used, the results of either or both of these measures may be added to the fingerprint matrix generated by the SVD operation.

FIG. 3 is a flow diagram of a process executed by the fingerprint analysis engine 20 for analyzing an extracted audio fingerprint for a match against registered fingerprints according to one embodiment of the invention. The process starts, and in step 200, the fingerprint analysis engine 20 receives a fingerprint (fingerprint matrix X) of an audio piece to be identified from the fingerprint extraction engine 18. The fingerprint analysis engine 20 then invokes a search and retrieval routine on the fingerprint database 26 with fingerprint matrix X. In this regard, the fingerprint analysis engine 20 inquires in step 202 whether there are more fingerprints in the fingerprint database 26 to compare. If the answer is NO, then all of the fingerprints in the database have been analyzed without finding a match. In this scenario, the fingerprint analysis engine returns a no match result in step 204.

On the other hand, if there are more fingerprints in the fingerprint database that have not been analyzed, the fingerprint analysis engine 20 computes in step 206, a difference between the fingerprint matrix X and a current fingerprint (fingerprint matrix Y) in the fingerprint database 26. According to one embodiment of the invention, the difference is computed by taking the well-known Euclidian

distance measure D for each row vector of the fingerprint matrices X and Y as follows:

$$D = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_m - y_m)^2}$$

where x_1, x_2, \dots, x_m are the values of a row vector of fingerprint matrix X, and y_1, y_2, \dots, y_m are the values of a row vector of fingerprint matrix Y. The distance measures for all the rows of the matrices are summed and, according to one embodiment of the invention, normalized. In step 208, a determination is made as to whether the sum of the distances exceed a threshold value. If the answer is NO, a match is declared. Otherwise, a next fingerprint in the fingerprint database is examined for a match.

According to one embodiment of the invention, if prior art fingerprinting techniques are also introduced, the time-indexed vectors generated by these techniques are measured for distance against corresponding stored fingerprint vectors and scaled by an appropriate constant. The resulting distance calculation is added to the distance calculation computed in step 206. A weighing factor may also be introduced to give more or less weight to the distance calculation performed by a particular technique. The total distance computation is then tested against the threshold value to determine if a match has been made.

FIG. 4 is a flow diagram of a process executed by the fingerprint analysis engine 20 for analyzing the extracted audio fingerprint for a match against registered fingerprints according to an alternative embodiment of the invention.

According to this embodiment, the process starts, and in step 300, the fingerprint analysis engine 20 receives the fingerprint (fingerprint matrix X) of the audio piece to the identified from the fingerprint extraction engine 18. The fingerprint analysis engine 20 invokes the indexing engine 22 in step 302 to identify the index of a subset of fingerprints in the fingerprint database 26 that, if a candidate matching the extracted fingerprint exists, contains the candidate. In this regard, the indexing engine 22 generates a query index for the extracted fingerprint. According to one embodiment of the invention, the index consists of four unordered numbers, and a match is deemed to have been made if an index exists in the fingerprint database that has three identical numbers, in any order, as that of the query index.

The remainder of the process of FIG. 4 continues in the same manner as in FIG. 3, except that the search space is limited to the subset of fingerprints identified by the matching index.

In this regard, the fingerprint analysis engine 20 inquires in step 304 whether there are more fingerprints in the identified subset of the fingerprint database 26 to compare. If the answer is NO, the fingerprint analysis engine returns a no match result in step 306.

If there are more fingerprints in the subset that have not been analyzed, the fingerprint analysis engine 20 computes in step 308 a difference between fingerprint matrix X and a current fingerprint (fingerprint matrix Y) in the subset. In step 310, a determination is made as to whether the difference exceeds a threshold value. If the answer is NO, a match is

declared. Otherwise, a next fingerprint in the identified subset is examined for a match.

5 FIG. 5 is a flow diagram of a process executed by the indexing engine 22 for assigning a database index to an audio piece according to one embodiment of the invention. The database index is used to identify a subset of fingerprints in the fingerprint database 26 for registering a fingerprint
10 extracted by the fingerprint extraction engine 18, or for reducing the candidates that need to be examined in the fingerprint database 26 for a match against the extracted fingerprint.

15 The process illustrated in FIG. 5 starts, and in step 400, either the indexing engine 22 or a separate fourier transform engine (not shown) calculates the FFT or the like of the audio piece preprocessed by the preprocessor 16 and obtains an FFT spectrum of the audio piece. In step 402, the
20 indexing engine 22 automatically obtains a list of notes of the audio piece. The list of notes are obtained via any of the well-known peak-tracking algorithms that exist in the prior art.

25 The peak-tracking algorithm generates tracks of local peaks in the FFT which are then analyzed by the indexing engine for their prominency. In this regard, the indexing engine 22 determines in step 404 whether there are any more
30 tracks to examine. If the answer is YES, the engine converts, in step 406, the track's frequency into an integer value that quantizes the track's frequency. According to one embodiment of the invention, this is done by quantizing the track's
35 frequency to a closest MIDI (Musical Instrument Digital

Interface) note number in a manner that is well known in the art.

In step 408, the indexing engine 22 computes a prominence value for the track based on factors such as, for example, the track's strength and duration. In step 410, the engine associates the computed prominence value to the track's MIDI note. In step 412, the prominence value for the MIDI note is accumulated into a prominence array. The process then returns to step 404 for analyzing a next track.

If there are no more tracks to examine, the indexing engine 22 selects in step 414, the MIDI note numbers in the prominence array with the highest prominence values and outputs them as an index of the associated subset in the fingerprint database 26. According to one embodiment of the invention, the four MIDI note numbers with the highest prominence values are selected for the index. According to one embodiment of the invention, the index consists of four unordered numbers where the numbers are the selected MIDI note numbers, rendering a total of 24 possible combinations for the index.

FIG. 6 is a flow diagram of a process for generating an identifier for a particular musical class according to one embodiment of the invention. Although this diagram is described in terms of the musical class, a person of skill in the art should recognize that the process extends to all types of audio and audio classes that may be conventional in the art.

The process starts, and in step 500, a set of audio pieces that belong to the musical class are selected. The selection of the pieces may be manual or automatic.

1 In step 502, the class identification engine computes a
 fingerprint for each audio piece in the set. According to one
 5 embodiment of the invention, the class identification engine
 invokes the fingerprint extraction engine 18 to compute the
 fingerprints via SVD operations. Other fingerprinting
 mechanisms may also be used in lieu and/or addition of the SVD
 10 fingerprinting mechanism.

10 In step 504, the class identification engine 24
 calculates an average of the fingerprints generated for the
 set. In this regard, the class identification engine computes
 a matrix, referred to as a class ID matrix, that minimizes a
 15 distance measure to all the audio pieces in the set in a
 manner that is well known in the art.

 In step 506, the calculated average of the fingerprints
 represented by the class ID matrix is associated with the
 20 musical class and in step 508, stored in the musical class
 database 28 as its identifier along with other information
 about the musical class. Such additional information may
 include, for example, a list of audio pieces that belong to
 the class, links to the fingerprint database 26 of audio
 25 fingerprints of the audio pieces that belong to the class,
 links to the audio profile vectors for the audio pieces that
 belong to the class, and/or the like.

 Once the identifiers for the musical classes have been
 30 generated, calculations may be made to determine how close or
 far an audio piece is to a particular musical class. This may
 be done, for example, by computing the distance between the
 fingerprint extracted for the audio piece and the class ID
 35 matrix for the particular musical class.

According to one embodiment of the invention, the various musical classes are used as attributes of an audio piece's audio profile vector. The distance calculations are stored in the audio profile vector for each attribute as an indication of how close the audio piece is to the associated musical class.

 FIG. 7 is a schematic block diagram of a computer network with one or more devices utilizing the audio fingerprinting system 10 of FIG. 1 according to one embodiment of the invention. The network includes a server 600 coupled to one or more end terminals 602-608 over a public or private network such as, for example, the internet 610. The end terminals may take the form of personal computers 602, personal digital assistants 604, laptops 606, wireless devices 608, and/or other types of stationary or mobile terminals that are known in the art.

 According to one embodiment of the invention, the audio fingerprinting system 10 resides in the server 600. Portions of the audio fingerprinting system may also reside in end terminals 602-608. The server 600 and/or end-terminals 602-608 may also include the music profiler disclosed in U.S. Patent Application Ser. No. 10/278,636, for automatically analyzing an audio piece and generating an audio profile vector. One or more processors included in the server 600 and/or end terminals 602-608 may further be configured with additional functionality to recommend audio pieces to users based on their preferences. Such functionality includes generating/retrieving audio profile vectors quantifying a plurality of attributes associated with the audio pieces in the audio database, generating/updating user preference

vectors, and selecting audio pieces from the audio database based on the user profile vector.

5 In an exemplary usage of the fingerprinting system 10, a user rates a song that does not have descriptive information associated with it. Instead of transmitting the entire song that the user wants to rate, a fingerprint of the song is transmitted along with the rating information. In this
10 regard, an end terminal used by the user accesses the server 600 and downloads an instance of the fingerprint extraction engine 18 into its memory (not shown). The downloaded fingerprint extraction engine 18 is invoked to extract the
15 fingerprint of the audio piece that is being rated. The extracted fingerprint is transmitted to the server 600 over the internet 610.

 Upon receipt of the extracted audio fingerprint, the
20 server 600 invokes the fingerprint analysis engine 20 to determine whether the received fingerprint is registered in the fingerprint database 26. If a match is made, the server retrieves the audio profile vector associated with the
25 fingerprint and uses it to update or generate a user profile vector for the user as is described in further detail in U.S. Patent Application Ser. No. 10/278,636. The user profile vector is then used to recommend other songs to the user.

30 If a match may not be made, the audio piece is analyzed, preferably by the end terminal, for generating the audio profile vector as is disclosed in further detail in U.S. Patent Application Ser. No. 10/278,636.

35 According to one embodiment of the invention, the end terminal may also download an instance of the indexing engine 22 for determining the index of the subset of fingerprints to

which the audio piece that is being rated belongs. The indexing information is then also transmitted to the server 600 along with the fingerprint information to expedite the search of the fingerprint database 26.

Although this invention has been described in certain specific embodiments, those skilled in the art will have no difficulty devising variations to the described embodiment which in no way depart from the scope and spirit of the present invention. Moreover, to those skilled in the various arts, the invention itself herein will suggest solutions to other tasks and adaptations for other applications.

For example, the audio fingerprinting system 10 may have applications above and beyond the recognition of audio pieces for generating audio profile vectors. For example, the system 10 may be used to find associated descriptive data (metadata) for unknown pieces of music. The system 10 may also be used to identify and protocol transmitted audio program material on broadcasting stations for verification of scheduled transmission of advertisement spots, securing a composer's royalties for broadcast material, or statistical analysis of program material.

It is the applicants intention to cover by claims all such uses of the invention and those changes and modifications which could be made to the embodiments of the invention herein chosen for the purpose of disclosure without departing from the spirit and scope of the invention. Thus, the present embodiments of the invention should be considered in all respects as illustrative and not restrictive, the scope of the invention to be indicated by the appended claims and their equivalents rather than the foregoing description.